

2010. 05.29

Bayesian Inference: with ecological applications 輪読会

担当: 深谷肇一(北海道大学環境科学院)

# Chapter 7

## Multimodel Inference

### (後半)

## 7.3 MULTIMODEL COMPUTATION

### □ 計算によるBMIは難しい。

- 通常の事後分布+ベイズファクターの計算。

### □ ベイズファクターを直接計算するために、各モデルの周辺尤度 $[Y|M]$ を計算したいが、ほとんどの場合不可能。

→ 代わりとなるアプローチが2つ。

### 1. モデルを非観測の量としてギブスサンプリングを行う。

1-1. BUGSを用いた例\*<sup>1</sup> (Section 7.3.1)

1-2. Reversible Jump MCMC\*<sup>2</sup> (Section 7.3.2)

### 2. BICを用いて $[Y|M]$ を近似する(Section 7.3.3)。

(\*<sup>1,2</sup> これらはともにRJCMCMCの例題になっている?)

## 7.3 MULTIMODEL COMPUTATION

### 7.3.1 Multimodel Inference in BUGS

モデルをカテゴリカルな確率変数として、BUGSで実装できる。

注意点： **モデルの自己相関**に気をつける。

#### *An Example with Nonnested Models*

#### □ 幾何分布モデルとポアソン分布の比較(Section 7.1.2)

- モデルはカテゴリカル変数。
  - 幾何分布モデル = 1、ポアソン分布モデル = 2
- モデルの事前確率は(0.1:0.9)に設定。
- ベイズファクターは「事後モデルオッズ / 事前モデルオッズ」として計算。

## 7.3 MULTIMODEL COMPUTATION

### Panel 7.1 BUGSコード

```
model {
  Model ~ dcat(pi[1:2])
  mu ~ dunif(0, 1000)
  lambda <- mu           # Poisson平均
  p <- 1 / (1 + mu)     # 幾何分布の成功確率

  for (i in 1:5) {
    GeomProb[i] <- p * pow(1 - p, Y[i])
    PoisProb[i] <- exp(-lambda + Y[i] * log(lambda) - logfact(Y[i]))

    b[i] <- pow(GeomProb[i], equals(Model, 1))
      + pow(PoisProb[i], equals(Model, 2)) - 1
                                                    # Model = 1なら幾何分布の尤度
    B[i] <- 1
                                                    # Model = 2ならポアソン分布の尤度
    B[i] ~ dbern(b[i]) # "1's trick"
  }
}
```

#### ■ “1's trick”

→ ベルヌーイ分布に従う確率変数 $B_i$  (成功確率 =  $b_i$ ) = 1と仮定することで、 $b_i$  を尤度と定義 ( $\Pr(Y_i = y_i) = b_i$ )。

## 7.3 MULTIMODEL COMPUTATION

### □ 結果

- 長さ100万の鎖を生成。
- モデル1:モデル2 = 605858:394142

$$\rightarrow BF_{1,2} = \left( \frac{605858}{394142} \right) / \left( \frac{0.1}{0.9} \right) = 13.83$$

- ・ 解析解は13.84(Section 7.1.2)。
- ・ ベイズファクターは事後モデル確率がほぼ等しい時に最も正確に推定されるので、事前分布を変えながら何回か推定するのがよい(ベイズファクターは事前モデルウェイトに依存しないので、この方法に不正はない)。
- Fig. 7.5:  $\mu$  の事後分布。
  - ・ ミューのMCMC標本平均は4.22。
    - ← モデルウェイトに基づく $\mu$ の加重平均になっている。

### *An Example with Nested Models*

#### □ 標識個体の回収率モデル(Section 11.2)

##### ■ 3年データ

- ・ 毎年はじめに400羽の鳥を放す。
  - ・ データ $X_{i,j}$ :  $i$ 年に放されて、 $j$ 年のはじめまでに死んでいるのが見つかった個体数。  
( $i = 1, 2, 3, j = i, i + 1, \dots, 3$ )
  - ・ データ $X_{i,4}$ :  $i$ 年に放されて、最後まで見つからなかった個体数。
  - ・  $f_i$ :  $i$ 年の回収率。
  - ・  $S_i$ :  $i$ 年の生存率。
- $X_{i,j}$ は多項分布(生起確率は $f_i$ 、 $S_i$ が制御)に従う。

## 7.3 MULTIMODEL COMPUTATION

### □ 4つのモデルを考える

- モデル1: 回収率、生存率ともに年により異なる。
- モデル2: 回収率は一定。
- モデル3: 生存率は一定。
- モデル4: 回収率、生存率ともに一定。

### □ モデル2, 3, 4はモデル1にネストしている。

→ モデルを表すカテゴリカル変数によって制御される”**パラメータのパレット**”を考えることで、モデルを包括的に記述する。

## 7.3 MULTIMODEL COMPUTATION

### Panel 7.2 BUGSコード

```
model {
  for (i in 1:3) {
    xf[i] ~ dunif(0, 1)
    f[i] <- pow(xf[1], constF) * pow(xf[i], 1 - constF)
    p[i, 4] <- 1 - sum(p[i, i:3])
    X[i, i:4] ~ dmulti(p[i, i:4], 400)
  }
  for (i in 1:2) {
    xS[i] ~ dunif(0, 1)
    S[i] <- pow(xS[1], constS) * pow(xS[i], 1 - constS)
  }
  Model ~ dcat(pi[1:4]) # モデルを表すカテゴリ変数
  constF <- equals(Model, 2) + equals(Model, 4)
  constF <- equals(Model, 3) + equals(Model, 4)
  p[1, 1] <- f[1]
  p[1, 2] <- S[1] * f[2]
  p[1, 3] <- S[1] * S[2] * f[3] # p[i, j]: X[i, j]の生起確率
  p[2, 2] <- f[2]
  p[2, 3] <- S[2] * f[3]
  p[3, 3] <- f[3]
}
# xf[i]とxS[i]がパラメータパレット
# (Modelに応じてf[i]とS[i]を適切に決定する)
```

## 7.3 MULTIMODEL COMPUTATION

□ ネストモデルはパラメーターの数がモデルによって異なる。

→ Reversible Jump MCMCを使わないと混交が悪くなる。

■ モデル $M_i$ には含まれないが、モデル $M_j$ には含まれるパラメーター $\theta_j$ がある場合( $i \neq j$ , Fig. 7.6参照):

- ギブスサンプリングのアルゴリズムでは、Model =  $M_i$  の場合でも $\theta_j$ をサンプルしないといけませんが、 $M_i$ は $\theta_j$ についての情報を一切持たないので、 $\theta_j$ は事前分布からそのままサンプルされる。
- $\theta_j$ の事前分布がvagueな場合、サンプルされた $\theta_j$ の値は大きくばらつくため、時々尤度の低い値に”迷い込む”。
- すると、次のステップでモデルを表す変数を選ぶとき、 $M_j$ はあまり”魅力的”でなくなる(選ばれる確率が低くなる)。

### 7.3.2 Reversible Jump Markov Chain Monte Carlo

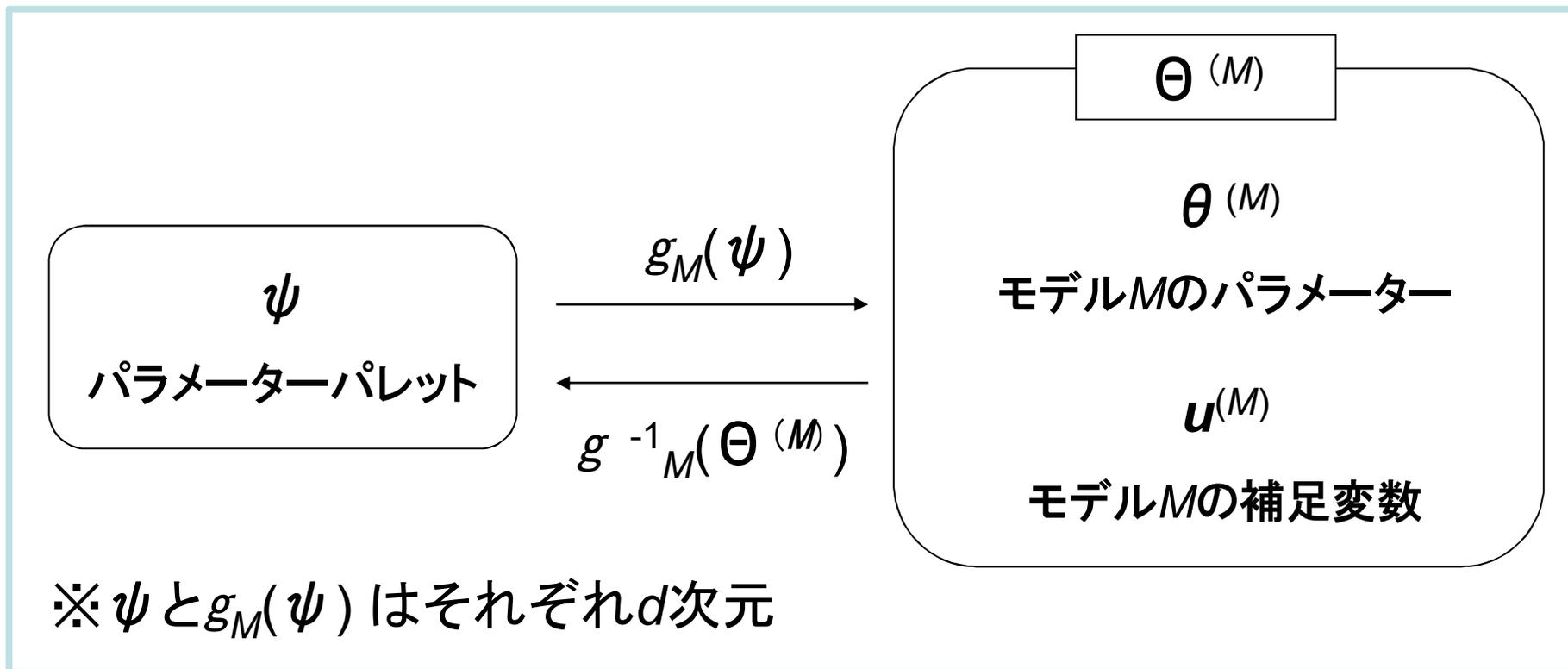
#### □ BMIを行うための効果的なMCMC実装方法。

- 全てのモデルのパラメーターを同時に制御する、“パラメーターパレット $\psi$ ”を利用。
- モデル $M$ と $\psi$ を交互に更新するギブスサンプリング
  - ・ パラメーターパレット  $\psi$ 
    - 最複雑モデルと同じか、より高次の次元 $d$ のベクトル。
  - ・ 写像  $g_M(\cdot) = g_M(\psi) = (\theta^{(M)}, u^{(M)}) = \Theta^{(M)}$ 
    - $g_M(\cdot)$ を $\psi$ に適用すると、モデル $M$ のパラメーター $\theta^{(M)}$ と、補足的な変数 $u^{(M)}$ を得る。
  - ・ 補足的変数  $u$ 
    - “がらくた”。 $\psi$ と $g^{(M)}(\psi)$ の次元を合わせるために用いる。

### 7.3 MULTIMODEL COMPUTATION

□  $g_M(\cdot)$ は可逆的。

→  $\theta^{(M)}$ と $u^{(M)}$ が得られれば、 $g_M^{-1}(\Theta^{(M)})$ より $\psi$ の唯一の対応値が得られる。



## 7.3 MULTIMODEL COMPUTATION

### □ 実例(Panel 7.2)

■ パラメーターパレット( $\psi$ ): `xf[1]`, `xf[2]`, `xf[3]`, `xS[1]`, `xS[2]`

■  $g_M(\psi)$ : `xf[i] -> f[i]` or `xf[1] -> f[i]`  
`xS[i] -> S[i]` or `xS[1] -> S[i]`

$$\boxed{M=2} \quad g_2(\Psi) = g_2 \begin{pmatrix} xf[1] \\ xf[2] \\ xf[3] \\ xS[1] \\ xS[2] \end{pmatrix} = \Theta^{(2)} = \begin{pmatrix} \theta^{(2)} \\ u^{(2)} \end{pmatrix} = \begin{pmatrix} f. \\ S_1 \\ S_2 \\ u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} xf[1] \\ xS[1] \\ xS[2] \\ xf[2] \\ xf[3] \end{pmatrix} \begin{array}{l} \theta^{(2)} \\ u^{(2)} \end{array}$$

$$\boxed{M=4} \quad g_4(\Psi) = g_4 \begin{pmatrix} xf[1] \\ xf[2] \\ xf[3] \\ xS[1] \\ xS[2] \end{pmatrix} = \Theta^{(4)} = \begin{pmatrix} \theta^{(4)} \\ u^{(4)} \end{pmatrix} = \begin{pmatrix} f. \\ S. \\ u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} xf[1] \\ xS[1] \\ xf[2] \\ xf[3] \\ xS[2] \end{pmatrix} \begin{array}{l} \theta^{(4)} \\ u^{(4)} \end{array}$$

### 7.3 MULTIMODEL COMPUTATION

#### □ パラメーターパレット $\psi$ のモデル特異的事前分布 $[\psi | M]^{\ast 3}$ の求め方

→ 同時分布  $f_M(\Theta^{(M)}) = [\theta^{(M)}, u^{(M)} | M]$  を指定し、変数変換の定理 (section 2.2.4) の多変量版を使う。

$$[\psi | M] = f_M(\underbrace{g_M(\psi)}_{\Theta^{(M)}}) \left| \mathbf{J}_{g_M}(\psi) \right|$$

$\mathbf{J}_{g_M}(\psi)$ :  $g_M(\psi)$  のヤコビ行列式

■  $[\theta^{(M)}, u^{(M)} | M]$  の事前分布なら、事前分布の独立性から簡単に求まる。

$$[\theta^{(M)}, u^{(M)} | M] = [\theta^{(M)} | M] \times [u^{(M)} | M]$$

(\*<sup>3</sup> モデル特有パラメーター  $[\theta^{(M)} | M]$  の事前分布と関連)

## 7.3 MULTIMODEL COMPUTATION

### *Gibbs Sampling*

□  $[\psi | M, Y]$ と $[M | \psi, Y]$ のfull conditional distributionsからの周期的なサンプリングによって構成される。

■  $\psi$ のサンプリング( $M$ は固定)

→  $\psi$ のfull-conditional

$$[\psi | \bullet] \propto [Y | \psi, M] \times [\psi | M]$$

からサンプリングする

or →  $\theta^{(M)}$ のfull-conditional

$$[\theta^{(M)} | \bullet] \propto [Y | \theta^{(M)}, M] \times [\theta^{(M)} | M]$$

からサンプリングして、逆変換  $\psi = g^{-1}_M((\theta^{(M)}, u^{(M)})')$  を行い、 $\psi$ を得る。

## 7.3 MULTIMODEL COMPUTATION

### ■ $M$ のサンプリング( $\psi$ は固定)

→ M-Hアルゴリズムを用いる。

- 現在の $M$ の値を所与として、候補値 $M^*$ が確率 $J(M^*|M)$ でランダムに選ばれる。
- ベルヌーイ試行によって、 $M$ か $M^*$ がfull conditional $[M|\cdot]$ からのサンプルとして抽出される。
- 候補値( $M^*$ )は確率 $\min(r, 1)$ によって選ばれる。

$$r = \frac{[Y|\psi, M^*][\psi|M^*]}{[Y|\psi, M][\psi|M]} \times \frac{[M^*]J(M|M^*)}{[M]J(M^*|M)}$$

## *RJMCMC for Comparing Two Binomial Models*

- 独立な二項変数  $Y_i \sim B(N_i, p_i)$ ,  $i = 1, 2$  について考える。
  - データ  $\rightarrow (Y_1, Y_2)$
  - モデル  $\rightarrow$  モデル1:  $p_1 \neq p_2$   
 モデル2:  $p_1 = p_2 \equiv \pi$
  - 未知変数  $\rightarrow M$ : モデルを表す確率変数  
 $\psi$ :  $p_1, p_2, \pi$  を決定するパラメーター(2次元)
  - 写像  $g(\cdot)$   $\rightarrow g_1(\psi) = (p_1, p_2)'$   
 $g_2(\psi) = \left(\frac{\psi_1 + \psi_2}{2}, \psi_2\right)' = (\pi, u)'$
  - 事前分布
    - $\rightarrow$  モデル1:  $p_i$  ( $i = 1, 2$ ) についてそれぞれ独立な  $\text{Be}(a_i, b_i)$
    - モデル2:  $\pi$  について  $\text{Be}(a_\pi, b_\pi)$ 、 $u$  について  $\text{Be}(a_u, b_u)$

## 7.3 MULTIMODEL COMPUTATION

### ■ $\psi$ の更新( $M$ は固定)

→ モデル1:  $\psi_i \sim Be(Y_i + a_i, N_i - Y_i + b_i)$

モデル2:  $\pi \sim Be(Y_1 + Y_2 + a_\pi, N_1 + N_2 - Y_1 - Y_2 + b_\pi)$

$u \sim Be(a_u, b_u)$

$\psi = g^{-1}_2(\Theta^{(2)})$  より  $\rightarrow \psi_1 = 2\pi - u, \psi_2 = u$

### ■ $M$ の更新( $\psi$ は固定)

1. まず式(7.14)を単純化。

- 事前モデルウェイトを等しくする  $\rightarrow [M]$ と $[M^*]$ をキャンセル。
- 現在のモデルを候補としない  $\rightarrow J(M|M^*)$ と $J(M^*|M)$ をキャンセル。

$\rightarrow [Y|\psi, M^*], [\psi|M^*], [Y|\psi, M], [\psi|M]$ だけ必要。

## ■ $M$ の更新(続き)

2. ( $[Y|\psi, M]$ はモデルの指定時に決まっているので) $[\psi|M]$ を求めれば $r$ が計算できる。

- Model = 1 のとき:  $g_1$  のヤコビ行列は  $2 \times 2$  の正方行列で、固有値は1。

$$\rightarrow [\Psi | M_1] = \prod_{i=1}^2 \frac{\Gamma(a_i + b_i)}{\Gamma(a_i)\Gamma(b_i)} \psi_i^{a_i-1} (1 - \psi_i)^{b_i-1} \mathbf{I}(\psi_i \in (0,1))$$

- Model = 2 のとき:  $g_2$  のヤコビ行列は

$$\begin{pmatrix} 1/2 & 1/2 \\ 0 & 1 \end{pmatrix} \quad \text{で、固有値は} 1/2.$$

$$\rightarrow [\Psi | M_2] = \frac{\Gamma(a_\pi + b_\pi)}{2\Gamma(a_\pi)\Gamma(b_\pi)} \bar{\psi}^{a_\pi-1} (1 - \bar{\psi})^{b_\pi-1} \times \mathbf{I}(\bar{\psi} \in (0,1)) \mathbf{I}(\psi_2 \in (0,1))$$

## 7.3 MULTIMODEL COMPUTATION

### □ 特定のデータセットで考えてみる。

→ データ:  $Y_1 = 8, Y_2 = 16, N_1 = 20, N_2 = 30$

ハイパーパラメーター: すべて1(一様分布)

### ■ $\psi$ の更新

→  $M = 1$ :  $\psi_1 \sim \text{Be}(9, 13), \psi_2 \sim \text{Be}(17, 15)$

$M = 2$ :  $\pi \sim \text{Be}(25, 27), u \sim \text{Be}(1, 1)$

( →  $\psi_1 = 2\pi - u, \psi_2 = u$  )

### ■ $M$ の更新

・ モデル1 → モデル2:  $\min\{r, 1\}$

・ モデル2 → モデル1:  $\min\{1/r, 1\}$

$$r = \frac{\bar{\psi}^{24} (1 - \bar{\psi})^{26} (1/2) \mathbf{I}(\bar{\psi} \in (0,1)) I(\psi_2 \in (0,1))}{\psi_1^8 (1 - \psi_1)^{12} \psi_2^{16} (1 - \psi_2)^{14} \mathbf{I}(\psi_1 \in (0,1)) I(\psi_2 \in (0,1))}$$

## 7.3 MULTIMODEL COMPUTATION

### □ 計算結果

■ 長さ $10^7$ の鎖を生成。

→ モデル2の事後モデルウェイトが65.81%\*4

### □ 事前分布を適切に設定すると、 $M$ を有効にサンプリングできる。

■  $M = 2$ のとき、 $\psi_1$ が負になることがあり、この場合 $1/r$ が0になるのでモデル2→モデル1の推移が起こらなくなる。

→ この現象が起こる頻度は、モデル2における $u$ の事前分布に依存する。\*5

(\*4 解析解は65.80%)

(\*5  $\psi_1 = 2\pi - u$ より、 $u$ が大きくならないようにすることで問題が起こりにくくなると思われる。)

## 7.3 MULTIMODEL COMPUTATION

### 7.3.3 Bayesian Information Criterion

□ ベイズファクター(周辺モデル尤度 $[Y|M]$ の比)を最尤推定に基づいて近似する。

→ cf. Section 7.3.1-2: ベイズファクターの計算にMCMCを応用。

■ BICは $-2\log([Y|M])$ の近似値。

$$BIC_M = -2\log(\mathbf{Y} | \hat{\theta}_M, M) + k \log(n) \approx -2\log(\mathbf{Y} | M)$$

・ 仮定:

1. データベクトル $Y$ は $n$ 個の観察からなり、モデル $M$ と $k$ 次元のパラメーター $\theta_M$ 所与のもとで条件付き独立。
2. 事後分布 $[\theta_M | M, Y]$ は多変量正規分布によって厳密に近似される。<sup>\*6</sup>
3.  $\theta_M$ の事前分布は、“unit information prior”と呼ばれる多変量正規分布によって厳密に近似される。この分布は平均が $\theta_M$ の最尤推定値に等しく、その共分散行列は事前分布が標準的な一つのデータが持っている情報と同量の情報を持った、弱い事前情報を持っている。

(<sup>\*6</sup>比較的大きなデータセットの解析のときに妥当となる(ベイズ中心極限定理)。)

## 7.3 MULTIMODEL COMPUTATION

- 仮定が満たされているとき、ベイズファクターは

$$BF_{i,j} = \exp\left(-\frac{1}{2}(BIC_i - BIC_j)\right)$$

と近似される。

→ この近似は、**一回に一つのモデルをあてはめた場合**に有効。

- ・ **注意点**:このアプローチは**事前分布に規定**(仮定3)がある。

→ 異なる事前分布(特にinformative prior)を与えた場合のMMIの結果と異なる場合がある。

## 7.4 INDICES OF MODEL ACCEPTABILITY: AIC AND DIC

□ モデル許容度の理想的な指標はモデル尤度\*<sup>7</sup>

→ モデル尤度に基づくBMIは、MMIのgold-standardとなる方法\*<sup>8</sup>

■ しかし、BMIは計算上の難しさがあるので、PBFやBICといった指標の代わりとなるモデル許容度の代替指標が使いたい場合もあるかもしれない。

→ **AIC**(と、そのベイズ版アナログ**DIC**)

→ BMIとの関連は弱いですが、ad hocにモデルウェイトを計算するのに使われることがある。

(\*<sup>7,8</sup> 著者の考え。)

## 7.4 INDICES OF MODEL ACCEPTABILITY: AIC AND DIC

### 7.4.1 Akaike's Information Criterion

- AICは真のモデル $f$ から候補モデル $g$ へのKullback-Leibler(K-L)ダイバージェンスの指標を推定する。

$$K(f, g) = \int f(y) \log \left\{ \frac{f(y)}{g(y)} \right\} dy$$

→ 真の対数尤度と近似対数尤度( $\log(g(y))$ )の差の期待値

$$\begin{aligned} K(f, g) &= E_f \{ \log f(y) - \log g(y) \} \\ &= E_f \{ \log f(y) \} - E_f \{ \log g(y) \} \end{aligned} \quad (7.19)$$

- 第一項は候補モデル間で比較を行う際にはキャンセルされる。
- 第二項は $E_f \{ \log g(y) \}$ がわからないので計算できないが、 $f$ からの標本があるので、近似推定は可能。

## 7.4 INDICES OF MODEL ACCEPTABILITY: AIC AND DIC

□ 未知の $k$ 次元パラメーター  $\theta$  によってあらわされるモデル $M$ について考える。

■  $\hat{\theta}[M]$ がデータ $Y$ にもとづくMLEとすると、

$$\text{AIC} = -2 \log([ Y | \hat{\theta}_M M ]) + 2k$$

はモデル $M$ に対応したK-Lダイバージェンスの指標の最小値の推定量となる。

■ サンプルサイズが小さいときはAICcがよいとされている。

$$\text{AICc} = \text{AIC} + \frac{2k(k+1)}{n-k+1}$$

→ これらの値が小さいときはK-Lダイバージェンスが小さい証拠なので、モデルの当てはまりがよいことを表す。

## 7.4 INDICES OF MODEL ACCEPTABILITY: AIC AND DIC

- モデル選択のときに、 $\theta [M]$ のどの値を使うか？
  - BF、PBFは不確実性を表す分布から、**パラメーターの取りうる全ての値**を用いる。
    - ・ 事前分布がimplicitに設定された場合のBICも同様に解釈。
  - BIC、AIC、DICは**最尤推定値**を使う。
    - 罰則付き最大尤度

$$C(M) = \log( L(M, \hat{\theta}_M) ) - g(K_M)$$

- AICとBICは $-2C(M)$ という形で、 $g(K_M)$ はそれぞれ $k$ と $(2/k)\log(n)$

## 7.4 INDICES OF MODEL ACCEPTABILITY: AIC AND DIC

- AICはBICよりもパラメーターの多いモデルを好む。
  - BICの複雑性ペナルティはサンプルサイズにも依存
    - ・  $n \geq 8$  のとき、BICのペナルティはAICのそれより大きくなる。
  
- BICウェイトを用いてAICウェイトを導出することを考える。
  - BICを用いた近似モデルウェイト

$$[M_i | \mathbf{Y}] \approx \frac{\exp\left(-\frac{1}{2} \text{BIC}_i\right) \pi_i}{\sum_j \exp\left(-\frac{1}{2} \text{BIC}_j\right) \pi_j}$$

(7.20)

## 7.4 INDICES OF MODEL ACCEPTABILITY: AIC AND DIC

□ BICウェイトを用いてAICウェイトを導出することを考える。(つづき)

■ データに基づく事前モデルウェイト

$$\pi_i = \frac{\exp(k_i \ln(n) / 2 - k_i)}{\sum_j \exp(k_j \ln(n) / 2 - k_j)} \quad (7.21)$$

を式(7.20)に代入してAICウェイトを得る。

$$\omega_i \approx \frac{\exp\left(-\frac{1}{2} \text{AIC}_i\right)}{\sum_j \exp\left(-\frac{1}{2} \text{AIC}_j\right)}$$

## 7.4 INDICES OF MODEL ACCEPTABILITY: AIC AND DIC

- この導出はAICウェイトのベイズ的な正当化にあたりと言われている(Burnham & Anderson, 2002)。
- 疑問点：
  1. 導出にはBICによる近似が必要だが、BICそのものは厳密にはベイズ的ではない。
  2. 異例な事前モデルウェイト(式7.21)を使う必要がある。
    - 事前分布はデータがない状況での知識を表しているはずだが、式(7.21)は事前分布がサンプルサイズに依存することを表していて、おかしい。
- $n$ が大きくなればなるほど、事前モデルウェイトは複雑なモデルほど大きくなる(Figure 7.7)。

## 7.4 INDICES OF MODEL ACCEPTABILITY: AIC AND DIC

### 7.4.2 Deviance Information Criteria

- AICへの批判： 混合効果モデルのときにどう計算していいかわからない！
  - 有効パラメーター数 $k$ をどう計算すればよいのか？
    - ランダム効果が含まれている場合、 $K-L$ に基づくモデル選択の一般的なアプローチはわかりにくい(Burnham & Anderson, 2002)。
- **DIC**: ベイズモデリングのための情報基準量
  - 1. (BICと比較して)事前分布がほとんどなく、
  - 2. (AICと比較して)最尤推定によって当てはめられておらず、
  - 3.モデルがネストされておらず、階層モデルや時系列モデルのようにデータが独立でない場合といったベイズ的状况でも利用可能

## 7.4 INDICES OF MODEL ACCEPTABILITY: AIC AND DIC

□ **DIC**: ベイズモデリングのための情報基準量(つづき)

■ Deviance  $D(\theta) = -2\log[Y|M, \theta_M]$  とすると、DICはDevianceの事後平均値と、モデルの複雑さを表す  $p_D$  の和となる。

$$\text{DIC} = \bar{D}(\theta_M) + p_D$$

■  $p_D$ : モデルの有効パラメーター数\*<sup>9</sup>

$$p_D = \bar{D}(\theta_M) - D(\bar{\theta}_M)$$

$\bar{D}(\theta_M)$ : devianceの事後平均値

$D(\bar{\theta}_M)$ : パラメーターの事後平均によって計算された deviance\*<sup>10</sup>

(\*<sup>9</sup> vagueな事前分布が与えられた場合、 $p_D$ の代わりに  $p_V = \text{Var}(D(\theta)|Y)$  が推奨されている (Gelman et al. 2004)。)

(\*<sup>10</sup> 事後中央値や事後モードによって計算される場合もある。)

## 7.4 INDICES OF MODEL ACCEPTABILITY: AIC AND DIC

### 7.4.3 Example: Trout Return Rates

□ AIC, BIC, DICのパフォーマンスを比較する。

■ ブルンナー湖(NZ)産卵支流へのブラウントラウト回帰率

・ 1961個体が標識後に放流され、翌年94個体を再捕獲。

→ 回帰率が性別と体長に依存するか？

・  $Y_i$  : 個体 $i$ の回帰インジケータ。確率 $\pi$ のベルヌーイ乱数。

・  $S_i$  : 個体 $i$ がメスであることを表すインジケータ。

・  $L_i$  : 個体 $i$ の体長。

・ 5つの候補モデルを考える(Table 7.2)

→  $S_i, L_i, S_i L_i$ はそれぞれ標準化。

## 7.4 INDICES OF MODEL ACCEPTABILITY: AIC AND DIC

- ・ パラメーター  $\beta_k$  の事前分布： 平均0の正規分布。
- ・ 事前分布の精度は大域変数  $T$  によって制御される。<sup>\*11</sup>  
→  $k_j$  をモデル  $j$  における  $\beta_k$  の数とすると、事前分布の精度は

$$\text{Precision } (\beta_k | \text{Model } j) = \begin{cases} K_j T, & \beta_k \text{ in Model } j \\ T, & \text{otherwise} \end{cases} \quad (7.21)$$

理由1. モデルと独立に全パラメーターの事前分布の分散を割り当てると、簡単なモデルが有利になりやすい。

理由2. モデル  $j$  の中ですべての  $\beta_k$  の事前分布の分散が等しくなり、またモデル間で、事前分布の分散の全体的なレベルがすべて等しく  $1/T$  になる。

- ・  $T = 0.40$  または  $T \sim \text{Ga}(3.29, 7.80)$  として、 $\pi_j$  がおおよそ  $(0, 1)$  の一様分布となるようにする。
- ・ Panel 7.3: BUGSコード

(<sup>\*.11</sup> 式(7.22)は  $\beta_k$  がモデル  $j$  に含まれない場合にも  $\beta_k$  を定義する。この事前分布の指定は推定には影響を及ぼさないが、MCMCの実行に有効(モデル間の推移を促進))

## 7.4 INDICES OF MODEL ACCEPTABILITY: AIC AND DIC

### ■ 3通りの当てはめ方法を比較。

1. AICとBICを計算するために、最尤推定によって当てはめ。
2. DICとAitkin's PBFを計算するために、一つのモデルにつき一回のMCMCで当てはめ。
3. BUGSを用いたBMI。

→ 結果: Table 7.3。

- ・ AICとDIC、BICとBFはお互い似ている。

- 最尤推定が難しいとき、DICがAICの代替となりうる。また、最尤推定の実装が容易で事前分布がvagueなら、BICがBFの代替品となりうる。

- ・ AICはパラメーター数の多いモデルを好む。

- AIC・DICとBIC・BFの違いは、K-L事前分布(式7.21)の違い。

PBFはさらにこの傾向が強いが、これは別の理由(情報量の多い“事前分布”が用いられている)による。

## 7.4 INDICES OF MODEL ACCEPTABILITY: AIC AND DIC

- モデルの複雑さについての考え方は事前分布に明示的に反映させるべき。
- 複雑なモデルを重視したいなら、事前モデルウェイトを $\exp(k)$ に比例させ、最節約性を重視したいなら $\exp(-k)$ に比例させる。
  - Table 7.4: これらの事前モデルウェイトとK-L事前分布、一様分布の結果を比較。
  - Complex priorでもモデル1が選択される。複雑なモデルが選択されたのは偏見的なK-L priorのみ。

## 7.5 AFTERWORD

### *Use and Abuse*

- モデル選択とモデルアベレージング：少数の、明瞭に異なる、注意深く考えぬかれたモデルの選択を行うときに有効。
  - 非常に多くのモデルから選択するのには適していない。
    - データを激しく希釈してしまう。パラメーター数が多くなればなるほど、より多くのデータが必要。
  
- モデルアベレージング： **モデルの不確実性**を考慮
  - BMIはモデルの特定にデータが費やされてしまうので、推定と予測は多少正確さが下がってしまう。
  - しかし、一つのモデルでの推論は、そのモデルの条件付き推論であるということに気を付けるべき。

## 7.5 AFTERWORD

### *Model Averaging*

□ BMI: パラメーターパレット $\psi$ を用いた推論

■  $\psi$ のモデルについての周辺尤度を計算。

→ 推定値はモデル間で重み付け平均化されて評価。

■ 平均化した量は解釈可能か？

・ 環境収容力を仮定しない指数増加モデルと、有限の環境収容力を仮定するゴンペルツモデルの比較。

→ 環境収容力の平均を求めても意味不明。

・ スケールの違いも問題になる。

### *Complexity and AIC's Prejudice*

□ モデルの事前ウェイト：事前に簡単なモデルを優遇すべきか？

→ No。モデルが十分事実に適合している必要がある。

■ Fisher(1925a)の考え方：最初は簡単なモデルを好む

→ 複雑さを増加させるときは、データから強く支持された場合のみ。

■ 著者の考え方：節約性に対する立場をモデル事前分布に反映

・ K-L priorに基づくAICウェイトには反対。